Optimal Relay Selection with Channel Probing in Wireless Sensor Networks

K. P. Naveen and Anurag Kumar

Dept. of Electrical Communication Engg., Indian Institute of Science, Bangalore 560 012, India.

{naveenkp, anurag}@ece.iisc.ernet.in

Abstract

Motivated by the problem of distributed geographical packet forwarding in a wireless sensor network with sleep-wake cycling nodes, we propose a local forwarding model comprising a node that wishes to forward a packet towards a destination, and a set of next-hop relay nodes, each of which is associated with a "reward" that summarises the cost/benefit of forwarding the packet through that relay. The relays wake up at random times, at which instants they reveal only the probability distributions of their rewards (e.g., by revealing their locations). To determine a relay's exact reward, the forwarding node has to further probe the relay, incurring a probing cost. Thus, at each relay wake-up instant, the source, given a set of relay reward distributions, has to decide whether to stop (and forward the packet to an already probed relay), continue waiting for further relays to wake-up, or probe an unprobed relay. We formulate the problem as a Markov decision process, with the objective being to minimize the packet forwarding delay subject to a constraint on the effective reward (the difference between the total probing cost and the actual reward of the chosen relay). Our problem can be considered as a variant of the *asset selling problem* with partial revelation of offers.

The most general class of decision policies can keep awake any or all the relays that have woken up. In this paper, we study the optimum over a restricted class of policies which, at any time, can keep only one unprobed relay awake, in addition to the best among the probed relays. We prove that the optimum stopping policy over this class is of threshold type, where the same threshold is used at each relay wake-up instant. Numerically, we find that the performance of the optimum over the restricted class is very close to that over the unrestricted class.

I. Introduction

Consider a wireless sensor network deployed for the detection of a *rare event*, [1], [2], [3], e.g., forest fire, intrusion in border areas, etc. To conserve energy, the nodes in the network *sleep-wake* cycle whereby they alternate between an ON state and a low power OFF state. We are further interested in *asynchronous* sleep-wake where the point processes of wake-up instants of the nodes are mutually independent [1], [4].

In such networks, whenever an event is detected, an alarm packet (containing the event location and a time stamp) is generated and has to be forwarded, through multiple hops, to a control center (*sink*) where appropriate action could be taken. In the present work we consider the traffic model in which one alarm packet is generated for each detected event. One may argue that in the case of an emergency it may be better to simply broadcast the alarm packet. A naive uncontrolled broadcast can lead to *broadcast storm* [5]. There have been proposals for forwarding

packets via controlled broadcast, e.g., Gradient Broadcast (GRAB) [6]. It is, however, difficult to optimize the tradeoffs in these protocols, such as forwarding delay and network power consumption.

Instead, in our work we seek to optimize the forwarding of a single alarm packet, in the face of sleep-wake cycling. Natural extensions of this work would be to consider the controlled forwarding of multiple copies of the alarm. Also, instead of simple multihopping we could use cooperative forwarding techniques. We are considering these aspects in our ongoing work.

Now, since the network is sleep-wake cycling, a forwarding node (i.e., a node with an alarm packet) has to wait for its neighbors to wake-up before it can choose a neighbor as the next hop. Thus, there is a delay incurred, due to the sleep-wake process, at each hop enroute to the sink. We are interested in minimizing the total average delay subject to a constraint on some global metric of interest such as average hop counts, or the average total transmission power (sum of the transmission power used at each hop). Such a global problem can be considered as a stochastic shortest path problem [7], for which a distributed Bellman-Ford algorithm (e.g., the LOCAL-OPT algorithm proposed by Kim et al. in [1]) can be used to obtain the optimal solution. A major drawback with such an approach is that a pre-configuration phase is required to run such algorithms, which would involve exchange of several control messages.

The focus of our research is, instead, towards designing *simple forwarding rules* using only the *local information* available at a forwarding node. For instance, in our earlier work ([4], [8]) on the problem of minimizing total end-to-end delay subject to a hop-count constraint, we formulated a *local forwarding problem* of minimizing one hop delay subject to a constraint on the progress (towards the sink) made by the chosen relay. We showed that the tradeoff, between the end-to-end metrics (total delay and hop counts), obtained by applying the local solution at each hop enroute to the sink is comparable with that obtained by the distributed Bellman-Ford algorithm. However, in our earlier model we had not considered the effect of random channel gains on the relay selection process. Thus, the current work is an extension of our local forwarding problem where now, due to the channel gains, new features such as, the action to probe a relay (to learn its channel) and the associated energy cost (probing cost), have to be incorporated while choosing a next-hop relay.

Outline and Contributions: In Section II we formulate the local forwarding problem of choosing a relay for the next-hop in the presence of random sleep-wake cycling and unknown channel gains. Our problem is a variant of the optimal stopping problem where we have a probe action, in addition to the traditional stop and continue actions. The main technical contributions are:

- We characterize (in Section III) the optimal policy over a restricted class (referred to as RST-OPT) in terms of certain stopping and stopping/probing sets (Section IV).
- We prove that the stopping sets have threshold structure and further the thresholds are stage independent. This can be considered as a generalization of the one-step-look ahead rule (see the Remark following Theorem 2).
- Finally, through numerical work (Section VI) we find that the performance of RST-OPT is close to that of the global optimal (GLB-OPT).

For the ease of presentation, we have moved most of the proofs to the Appendix.

II. THE LOCAL FORWARDING PROBLEM: SYSTEM MODEL

In this section we will develop our system model from the context of *geographical forwarding* [4], [9], [10]. By geographical forwarding, also known as location aware routing, we mean that each node in the network knows its location (with respect to some reference) and the location of the sink.

Consider a forwarding node (henceforth referred to as *source*) at a distance v_0 from the *sink* (see Fig. 1). The *communication region* is the set of all locations where reliable exchange of *control messages* can take place between the source and a receiver, if any, at these locations. In Fig. 1 we have shown the communication region to be circular, but in reality this region can be arbitrary. The set of nodes within the communication region are referred to as *neighbors*. Let v_ℓ represent the distance of a location ℓ (a location is a point in $\Re^2$) from the sink. Then define the *progress* of the location ℓ as $Z_\ell := v_\ell - v_0$. The source is interested in forwarding the packet only to a neighbor within the *forwarding region* $\mathcal{L}$ where, $\mathcal{L} = \{\ell \in \text{communication region} : Z_\ell \geq 0\}$. The forwarding region is shown hatched in Fig. 1. We will refer to the nodes in the forwarding region as *relays*.

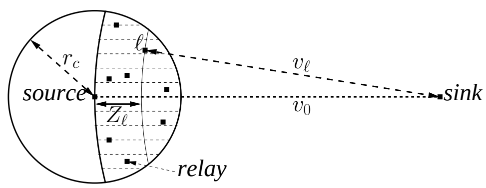


Fig. 1. The hatched area is the forwarding region $\mathcal{L}$. For a location $\ell \in \mathcal{L}$, the progress Z_ℓ is the difference between the source-sink distance, v_0, and the location-sink distance, v_ℓ.

For the channel between the source and a relay at location ℓ (if any) we will assume the Rayleigh fading model so that the received symbol Y at ℓ can be written as

$$Y_\ell = \sqrt{P d_\ell^{-\beta}} H_\ell X + W_\ell$$

where, P is the transmit power, d_ℓ is the distance between the source and the location ℓ, β is the path loss attenuation factor usually in the range 2 to 4, H_ℓ is the random channel gain ($H_\ell \sim \mathcal{C}(0,1)$), X is the unit-power transmitted symbol (i.e., $\mathbb{E}|X|^2 = 1$), W_ℓ is complex Gaussian noise of variance N_0 ($W_\ell \sim \mathcal{C}(0, N_0)$). By $H_\ell \sim \mathcal{C}(0,1)$, we mean that H_ℓ is complex Gaussian random variable of zero mean and unit variance.

We will assume that the set of channel gains, $\{H_\ell : \ell \in \mathcal{L}\}$, is *homogenous* i.e., $\{H_\ell\}$ are independent. We will further assume a *quasi static channel model* where, each time the source gets a packet to forward, the channel gains $\{H_\ell\}$ remain unchanged over the entire duration of the decision process. However, these would have changed the next time the source gets a packet to forward. In physical layer wireless terminology, we have a flat fading, slowly varying channel.

The instantaneous received SNR (signal to noise ratio) at ℓ is given by $SNR(P, d_\ell) = \frac{P d_\ell^{-\beta} |H_\ell|^2}{N_0}$. A transmission is considered to be successful (i.e., a receiver at ℓ can decode reliably) if the received $SNR(P, d_\ell)$ is above a

threshold Γ. Then, given the channel gain H_ℓ, the minimum power required to achieve the SNR constraint is $P_\ell = \frac{\Gamma N_0 d_\ell^\beta}{|H_\ell|^2}$.

Reward Structure: Several metrics have been proposed in the literature [11], to enable a forwarding node to evaluate its neighbors, before choosing one for the next hop. In the current work we prefer to use a reward metric, which is a combination of the progress (Z_ℓ) and the minimum power required (P_ℓ). Formally, to each location $\ell \in \mathcal{L}$ we associate a *reward* R_ℓ as

$$R_\ell = \frac{Z_\ell^a}{P_\ell^{(1-a)}} = \frac{Z_\ell^a}{(\Gamma N_0 d_\ell^\beta)^{(1-a)}}(|H_\ell|^2)^{(1-a)}, \tag{1}$$

where $a \in [0,1]$ is used to tradeoff between Z_ℓ and P_ℓ. The reward being inversely proportional to P_ℓ is clear, because it is advantageous to use less power to get the packet across. R_ℓ is made proportional to Z_ℓ to give a sense of progress towards the sink, while choosing a relay for the next hop. Further motivation for choosing the particular structure for the reward is available in Appendix VIII-A. Finally, let F_ℓ represent the distribution of R_ℓ. Thus, there is a collection of reward distributions $\mathcal{F}$ indexed by $\ell \in \mathcal{L}$. From (1), note that, given a location ℓ it is only possible to know the reward distribution F_ℓ. To know the exact reward R_ℓ, the source has to send probe packets and learn the channel gain H_ℓ.

Sleep-Wake Process: Without loss of generality, we will assume that the source gets a packet to forward (either from an upstream node or by detecting an event) at time 0. There are N relays in the forwarding set $\mathcal{L}$ that wake-up sequentially at the points of a renewal process, $W_1 \leq W_2 \leq \cdots \leq W_N$. Let $U_k := W_k - W_{k-1}$ ($U_1 := W_1$) denote the *inter-wake-up time* (renewal lifetime) between the k and the $(k-1)$-th relay. Then $U_1, U_2, \cdots, U_N$, are independent with their means, equal to τ. For example, U_k, $1 \leq k \leq N$, could be exponentially distributed with mean τ, or could be constant (deterministic) with value τ.

Sequential Decision Problem: Let $L_1, L_2, \cdots, L_N$, denote the relay locations which are i.i.d. uniform over the forwarding set $\mathcal{L}$. Let $A(\cdot)$ denote the uniform distribution over $\mathcal{L}$ so that, for $k = 1, 2, \cdots, N$, the distribution of L_k is A. The source (with a packet to forward at time 0) only knows that there are N relays in its forwarding set $\mathcal{L}$, but apriori does not know their locations (L_k) nor their channel gains (H_{L_k}). When the k-th relay wakes up, we assume that its location L_k and hence the reward distribution F_{L_k} is revealed to the source. This can be accomplished by including the location information L_k within a control packet (sent using low rate robust modulation technique, and hence, assumed to be error free) transmitted by the k-th relay upon wake-up. However, if the source wishes to learn the channel gain H_{L_k} (and hence the exact reward value R_{L_k}), it has to send additional probe packets incurring an energy cost of δ units. Thus, when the k-th relay wakes up (referred to as stage k) the actions available to the source are:

- stop and forward the packet to a previously probed relay accruing the reward of that relay. It is clear that it is optimal to forward, among all the probed relays, to the relay with the maximum reward. Thus, henceforth, the action stop always implies that the best relay (among those that have been probed) is chosen. With the stop action the decision process terminates.
- continue to wait for the next relay to wake-up (the average waiting time is τ) and reveal its reward distribution,

at which instant the decision process is regarded to have entered stage $k+1$.

- probe a relay from the set of all unprobed relays (provided there is at least one unprobed relay). The probed relay's reward value is revealed allowing the source to update maximum reward among all the probed relays. After probing, the decision process is still at stage k and again the source has to decide upon an action.

In the model, for the sake of analysis, we neglect the time taken to probe a relay and learn its channel gain. We also neglect the time taken for the exchange of control packets. This is reasonable for very low duty cycling networks where the average inter-wake-up times are much larger than the time taken for probing, or the exchange of control packets.

A *decision rule* or a *policy* is a mapping, at each stage, from all histories of the decision process to the set of actions. Let Π represent the class of all policies. For a policy $\pi \in \Pi$ the delay incurred, D_π, is the time until a relay is chosen (which is one of the W_k). Let R_π represent the reward of the chosen relay and M_π be the total number of relays that were probed during the decision process. Recalling that δ is the cost of probing, δM_π represents the total cost of probing using policy π. We would like to think of $(R_\pi - \delta M_\pi)$ as the *effective reward* obtained using policy π. The problem we are interested in is the following:

$$\begin{aligned} \min_{\pi \in \Pi} \quad & \mathbb{E}D_\pi \\ \text{Subject to:} \quad & \Big(\mathbb{E}R_\pi - \delta \mathbb{E}M_\pi\Big) \geq \gamma, \end{aligned} \tag{2}$$

where $\gamma > 0$ is a constraint on the effective reward. We introduce a Lagrange multiplier $\eta > 0$ and focus our attention towards solving the following unconstrained problem:

$$\min_{\pi \in \Pi} \Big(\mathbb{E}D_\pi - \eta \mathbb{E}R_\pi + \eta\delta\mathbb{E}M_\pi\Big). \tag{3}$$

We will call Π the *complete class* of policies, as Π contains policies that are allowed, for each stage k, to keep all the relays until stage k awake. Formally, if $b_k = \max\Big\{R_i : i \leq k, \text{ relay } i \text{ has been probed}\Big\}$ and $\mathcal{F}_k = \Big\{F_{L_i} : i \leq k, \text{ relay } i \text{ is unprobed}\Big\}$, then the decision at stage k is based on $(b_k, \mathcal{F}_k)$. Thus, the set of all possible states at stage k is large. Hence for analytical tractability we first consider (in Section III) solving the problem in (3) over a *restricted class* of policies $\overline{\Pi} \subseteq \Pi$ where a policy in $\overline{\Pi}$ is restricted to take decisions keeping only two relays awake, one the best among all probed relays and one among the unprobed ones, i.e., the decision at stage k is based on (b_k, G_k) where $G_k \in \mathcal{F}_k$.

A. Related Work

The work reported in this paper is an extension of our earlier work ([4], [8]). The major difference is that in [4] we assume that when a relay wakes up its exact reward value is revealed. This is reasonable if the reward is simply the geographical progress (towards the sink) made by a relay, which was the case in [4]. In [8] we studied a variant where the number of relay N is not known to the source. However, in [8] as well the exact reward valued is revealed by a relay upon wake-up.

There has been other work in the context of geographical forwarding and anycast routing, where the problem of choosing one among several neighboring nodes has been studied [10], [12]. The authors in [10] study the

policy of always choosing a neighbor that makes the maximum progress towards the sink. Thus, they do not study the tradeoff between the relay selection delay and the progress (or other reward metric), which is the major contribution of our work. For a sleep-wake cycling network, Kim et al. in [1] have developed a distributed Bellman-Ford algorithm (referred to as LOCAL-OPT) to minimize the average end-to-end delay. However a pre-configuration phase, involving lot of control message exchanges, is required to run the LOCAL-OPT algorithm.

In all of the above work, the metric that signifies the quality of a relay is always exactly revealed, and hence does not involve a probing action. The action to probe generally occurs in the problem of channel selection [13], [14]. We will discuss the parallels and differences of these works (particularly that of Chaporkar and Prouriere in [13]) with ours in detail in Section V.

Finally (but very importantly) our work can be considered as a variant of the *asset selling problem* which is a major class within the *optimal stopping problems* [15] (the other classes being the secretary problem, the bandit problem, etc.,). The basic asset selling problem [16], [17], comprises offers that arrive sequentially over time. The offers are i.i.d. As the offers arrive, the seller has to decide whether to take an offer or wait for future offers. The seller has to pay a cost to observe the next offer. Previous offers cannot be recalled. The decision process ends with the seller choosing an offer. Over the years, several variants of the basic problem have been studied, e.g., problems with uncertain recall [18], problems with unknown reward distribution [19], etc. In most of the variants, when an offer arrives, its value is exactly revealed to the seller, while in our case only the offer (i.e., reward) distribution is made available and the source, if it wishes, has to probe to known the exact offer value.

Close to our work is that of Stadje [20] where only some intial information about an offer (e.g., the average size of the offer) is revealed to the decision maker upon its arrival. In addition to the actions, stop and continue, the decision maker can also choose to obtain more information about the offer by incurring a cost. Recalling previous offers is not allowed. A similar problem is studied by Thejaswi et al. in [21], where initially a coarse estimate of the channel gain is made available to the transmitter. The transmitter can choose to probe the channel, the second time, to get better estimate. For both these works ([20], [21]), the optimal policies are characterized by thresholds. However, the horizon length of these problems is infinite, because of which the thresholds are stage independent. In general, for a finite horizon problem the optimal policy would be stage dependent. However, for our problem (despite being a finite horizon one) we are able to show that certain optimal stopping sets are identical at every stage. This is due to the fact that we are allowing the best probed relay to stay awake.

III. Restricted Class $\overline{\Pi}$: An MDP Formulation

In this section we will formulate our local forwarding problem as a Markov decision process [22], which will require us to first discuss, the one-step cost and the state transition.

A. One-Step Costs and State Transition

Recall that for any policy in the restricted class $\overline{\Pi}$, the decision at stage k should be based on (b_k, G_k) where b_k is the best reward so far and $G_k \in \mathcal{F}_k$ is the reward distribution of an unprobed relay that is retained until stage k.

If the source's decision is to stop, then the source enters a terminating stage ψ incurring a cost of $-\eta b_k$ (recall from (3) that η is the Lagrange multiplier).

If the action is to continue then the source will first incur a waiting cost of U_{k+1} (average cost is τ). When the $(k+1)$-th relay wakes-up (whose reward distribution is $F_{L_{k+1}}$), first the source has to choose between G_k and $F_{L_{k+1}}$, then put the other one to sleep so that the state at stage $k+1$ will again be of the form (b_{k+1}, G_{k+1}). Since the action at stage k was to continue, the set of probed relays at stage $k+1$ remain unchanged so that $b_{k+1} = b_k$.

Finally the source could, at stage k, probe the distribution G_k incurring a cost of $\eta\delta$. After probing the decision process is still considered to be at stage k where now the state is $b'_k = \max\{b_k, R_k\}$ where (R_k is a sample from the distribution G_k). The source has to now further decide whether to stop (incurring a cost of $-\eta b$ and enter ψ) or continue (cost is U_{k+1} and the next state is $(b'_k, F_{L_{k+1}})$). Note that for a policy π, the sum of all the one-step costs, starting from stage 1, will equal the total cost[1] in (3).

B. Bellman Equation and the Cost-to-go Functions

Let $J_k(\cdot)$, $k = 1, 2, \cdots, N$, represent the optimal cost-to-go function at stage k. Thus, $J_k(b)$ and $J_k(b, F_\ell)$ denote the cost-to-go, depending on whether there is or is not an unprobed relay. For the last stage, N, we have, $J_N(b) = -\eta b$, and

$$\begin{aligned} J_N(b, F_\ell) &= \min\Big\{ -\eta b, \eta\delta + \mathbb{E}_\ell\Big[J_N(\max\{b, R_\ell\})\Big]\Big\} \\ &= \min\Big\{ -\eta b, \eta\delta - \eta\mathbb{E}_\ell\Big[\max\{b, R_\ell\}\Big]\Big\} \end{aligned} \tag{4}$$

where $\mathbb{E}_\ell$ denotes expectation w.r.t. the distribution F_ℓ. The first term in the min-expression above is the cost of stopping and the second term is the average cost of probing and then stopping (the action continue is not available at the last stage N). Next, for stage $k = N-1, N-2, \cdots, 1$, denoting by $\mathbb{E}_A$ the expectation over the distribution A of the index, L_{k+1}, of the next relay to wake up, we have

$$J_k(b) = \min\Big\{ -\eta b, \tau + \mathbb{E}_A\Big[J_{k+1}(b, F_{L_{k+1}})\Big]\Big\} \tag{5}$$

and

$$J_k(b, F_\ell) = \min\Big\{ -\eta b, \eta\delta + \mathbb{E}_\ell\Big[J_k(\max\{b, R_\ell\})\Big], \tau + \mathbb{E}_A\Big[\min\{J_{k+1}(b, F_\ell), J_{k+1}(b, F_{L_{k+1}})\}\Big]\Big\}. \tag{6}$$

The last term in the min-expression of (5) and (6) is the average cost of continuing. When the state at stage k, $1 \le k \le N-1$, is (b, F_ℓ) and, if the source decides to continue, then the reward distribution of the next relay is $F_{L_{k+1}}$. Now, given the distributions F_ℓ and $F_{L_{k+1}}$, if the source is asked to retain one of them, it is always optimal to go with the distribution that fetches a lower cost-to-go from stage $k+1$ onwards, i.e., it is optimal to

[1] Since every policy has to invariably wait for the first relay to wake-up, at which instant the decision process begins, U_1 is not accounted for in the total cost by any policy π.

retain F_ℓ if $J_{k+1}(b,F_\ell) \leq J_{k+1}(b,F_{L_{k+1}})$, otherwise retain $F_{L_{k+1}}$[2]. In the next section we will show that if F_ℓ is *stochastically greater* than F_u then $J_{k+1}(b,F_\ell) \leq J_{k+1}(b,F_u)$ (see Lemma 2-(ii)).

First, for simplicity let us introduce the following notations. For $k = 1, 2, \cdots, N-1$, let $cc_k(\cdot)$ represent the cost of continuing, i.e.,

$$cc_k(b) = \tau + \mathbb{E}_A\Big[J_{k+1}(b, F_{L_{k+1}})\Big] \tag{7}$$

$$cc_k(b, F_\ell) = \tau + \mathbb{E}_A\Big[\min\{J_{k+1}(b,F_\ell), J_{k+1}(b,F_{L_{k+1}})\}\Big], \tag{8}$$

and, for $k = 1, 2, \cdots, N$, the cost of probing, $cp_k(\cdot)$, is

$$cp_k(b, F_\ell) = \eta\delta + \mathbb{E}_\ell\Big[J_k(\max\{b, R_\ell\})\Big]. \tag{9}$$

The above expressions will be used extensively from now on. From (7) and (8) it is immediately clear that $cc_k(b,F_\ell) \leq cc_k(b)$ for any F_ℓ ($\ell \in \mathcal{L}$). We note this down as a Lemma.

Lemma 1: For any b and F_ℓ we have $cc_k(b,F_\ell) \leq cc_k(b)$.

■

Using the above notation, the cost-to-go functions can be written as

$$J_k(b) = \min\Big\{-\eta b, cc_k(b)\Big\} \tag{10}$$

$$J_k(b, F_\ell) = \min\Big\{-\eta b, cp_k(b,F_\ell), cc_k(b,F_\ell)\Big\}. \tag{11}$$

C. Ordering Results for the Cost-to-go Functions

We begin with the definition of stochastic ordering.

Definition 1 (Stochastic Ordering): Given two distributions F_ℓ and F_u we say that F_ℓ is stochastically greater than F_u, denoted as $F_\ell \geq_{st} F_u$, if $F_\ell(x) \leq F_u(x)$ for all x. Alternatively, one could use the following definition: $F_\ell \geq_{st} F_u$ if and only if for every non-decreasing (non-increasing) function $f : \Re \to \Re$, $\mathbb{E}_\ell[f(R_\ell)] \geq \mathbb{E}_u[f(R_u)]$ ($\mathbb{E}_\ell[f(R_\ell)] \leq \mathbb{E}_u[f(R_u)]$) where the distributions of R_ℓ and R_u are F_ℓ and F_u, respectively. ■

The ordering properties of the cost-to-go functions are listed in the following lemma.

Lemma 2: For $1 \leq k \leq N$ (for part (iii), $1 \leq k \leq N-1$),

(i) $J_k(b)$ and $J_k(b,F_\ell)$ are decreasing in b.

(ii) if $F_\ell \geq_{st} F_u$ then $J_k(b,F_\ell) \leq J_k(b,F_u)$.

(iii) $J_k(b) \leq J_{k+1}(b)$ and $J_k(b,F_\ell) \leq J_{k+1}(b,F_\ell)$.

Proof: Part (i) and (iii) follow easily by straightforward induction. To prove part (ii), we need to use part (i) and the definition of stochastic ordering (Definition 1). Formal proof is available in Appendix VIII-B. ■

The following lemma is an immediate consequence of Lemma 2-(ii) and 2-(iii), respectively,

Lemma 3: (i) For $1 \leq k \leq N-1$, if $F_\ell \geq_{st} F_u$ then $cp_k(b,F_\ell) \leq cp_k(b,F_u)$ and $cc_k(b,F_\ell) \leq cc_k(b,F_u)$.

[2]Formally one has to introduce an intermediate state of the form $(b, F_\ell, F_{L_{k+1}})$ at stage $k+1$ where the only actions available are, choose F_ℓ or $F_{L_{k+1}}$. Then $J_{k+1}(b, F_\ell, F_{L_{k+1}}) = \min\{J_{k+1}(b,F_\ell), J_{k+1}(b,F_{L_{k+1}})\}$, which, for simplicity, we are directly using in (6).

(ii) For $1 \le k \le N-2$, $cc_k(b) \le cc_{k+1}(b)$, $cp_k(b, F_\ell) \le cp_{k+1}(b, F_\ell)$ and $cc_k(b, F_\ell) \le cc_{k+1}(b, F_\ell)$.

■

IV. RESTRICTED CLASS $\overline{\Pi}$: STRUCTURAL RESULTS

We begin by defining, at stage k, the *optimal stopping set* $\mathcal{S}_k$. Similarly, for a given distribution F_ℓ we define the *optimal stopping set* $\mathcal{S}_k^\ell$ and the *optimal stopping/probing set* $\mathcal{Q}_k^\ell$. For $k = 1, 2, \cdots, N-1$,

$$\mathcal{S}_k := \Big\{ b : -\eta b \le cc_k(b) \Big\}, \tag{12}$$

and, for a given distribution F_ℓ,

$$\mathcal{S}_k^\ell \quad := \quad \Big\{ b : -\eta b \le \min\{cp_k(b, F_\ell), cc_k(b, F_\ell)\} \Big\} \tag{13}$$

$$\mathcal{Q}_k^\ell \quad := \quad \Big\{ b : \min\{-\eta b, cp_k(b, F_\ell)\} \le cc_k(b, F_\ell) \Big\}. \tag{14}$$

From (5) it follows that the optimal stopping set $\mathcal{S}_k$ is the set of states b (states of these form are obtained after probing at stage k) where it is better to stop than to continue. Similarly from (6), the set $\mathcal{S}_k^\ell$ has to be interpreted as, for a given distribution F_ℓ, the set of b such that at the state (b, F_ℓ) it is better to stop, while the set $\mathcal{Q}_k^\ell$ is the set of b such that at (b, F_ℓ) it is better to either stop or probe.

From the definition of these sets, the following ordering properties are straightforward.

Lemma 4: For $k = 1, 2, \cdots, N-1$ and any F_ℓ we have (i) $\mathcal{S}_k^\ell \subseteq \mathcal{Q}_k^\ell$ (ii) $\mathcal{S}_k^\ell \subseteq \mathcal{S}_k$ (iii) if $F_\ell \ge_{st} F_u$ then $\mathcal{S}_k^\ell \subseteq \mathcal{S}_k^u$ and (iv) $\mathcal{S}_k \subseteq \mathcal{S}_{k+1}$ and $\mathcal{S}_k^\ell \subseteq \mathcal{S}_{k+1}^\ell$.

Proof: Recall the definitions in (12), (13) and (14). Then (i) simply follows from the definitions. For (ii), recall from Lemma 1 that $cc_k(b, F_\ell) \le cc_k(b)$. Finally, (iii) and (iv) are simple consequences of Lemma 3-(i) and 3(ii), respectively. ■

However it is not immediately clear how the sets $\mathcal{Q}_k^\ell$ and $\mathcal{S}_k$ are ordered. Later in this section we will, under the assumption that $\mathcal{F}$ is *totally stochastically ordered* (to be defined next), show that $\mathcal{S}_k \subseteq \mathcal{Q}_k^\ell$. This result will enable us to prove that the stopping sets, $\mathcal{S}_k$, are identical at every stage k (Theorem 2). Similarly, in Theorem 3 we will show that, for any distribution F_ℓ, $\mathcal{S}_k^\ell$ are also identical across all stages k. First we need the following assumption.

Assumption 1 (Total Stochastic Ordering Property of $\mathcal{F}$): From here on, we will assume that $\mathcal{F}$ is totally stochastically ordered meaning, any two distributions from $\mathcal{F}$ are always stochastically ordered. Formally, if $F_\ell, F_u \in \mathcal{F}$ then either $F_\ell \ge_{st} F_u$ or $F_u \ge_{st} F_\ell$. We further assume the existance of a *minimal distribution* $F_m \in \mathcal{F}$ such that for every $F_\ell \in \mathcal{F}$ we have $F_\ell \ge_{st} F_m$. ■

Note that such a restriction on $\mathcal{F}$ is not very stringent, in the sense that the set of distributions arising in our local forwarding problem in Section II, $\{F_\ell : \ell \in \text{forwarding set}\}$, is totally stochastically ordered.

A. Stopping Sets: Threshold Rule

Before showing the stage independence property of the stopping sets, first in this section we will prove another important property about the stopping sets, that they are characterized by thresholds. The following is the key lemma for proving such a property.

Lemma 5: For $1 \le k \le N-1$ and for $b_2 > b_1$ we have

(i) $cc_k(b_1) - cc_k(b_2) \le \eta(b_2 - b_1)$,

and for any distribution F_ℓ we have,

(ii) $cp_k(b_1, F_\ell) - cp_k(b_2, F_\ell) \le \eta(b_2 - b_1)$

(iii) $cc_k(b_1, F_\ell) - cc_k(b_2, F_\ell) \le \eta(b_2 - b_1)$.

Proof: See Appendix VIII-C. ■

Theorem 1: For $1 \le k \le N-1$ and for $b_2 > b_1$,

(i) if $b_1 \in \mathcal{S}_k$ then $b_2 \in \mathcal{S}_k$,

(ii) for any F_ℓ, if $b_1 \in \mathcal{S}_k^\ell$ then $b_2 \in \mathcal{S}_k^\ell$.

Proof: Part (i): Using part (i) of the previous lemma (Lemma 5-(i)) we can write,

$$-\eta b_2 \le -\eta b_1 - cc_k(b_1) + cc_k(b_2).$$

Since $b_1 \in \mathcal{S}_k$, from (12) we know that $-\eta b_1 \le cc_k(b_1)$, using which, in the above expression, we obtain $-\eta b_2 \le cc_k(b_2)$ so that $b_2 \in \mathcal{S}_k$. *Part (ii)* similarly follows using Lemma 5-(ii) and 5-(iii), respectively, to show $-\eta b_2 \le cp_k(b_2, F_\ell)$ and $-\eta b_2 \le cc_k(b_2, F_\ell)$. ■

Discussion: Thus, the stopping set $\mathcal{S}_k$ can be characterized in terms of a lower bound x_k as illustrated in Fig. 2(a). Similarly for distributions F_ℓ and F_u, there are (possibly different) thresholds x_k^ℓ and x_u^ℓ (Fig. 2(b) and 2(c)). Using Lemma 4-(ii) and 4-(iii), for $F_\ell \ge_{st} F_u$ we can write, $x_k \le x_k^u \le x_k^\ell$. These thresholds are for a given stage k. From Lemma 4-(iv), we know that the thresholds x_k and x_k^ℓ are decreasing with k. The main result in the next section (Theorem 2 and 3) is to show that these thresholds are, in fact, equal (i.e., $x_k = x_{k+1}$ and $x_k^\ell = x_{k+1}^\ell$).

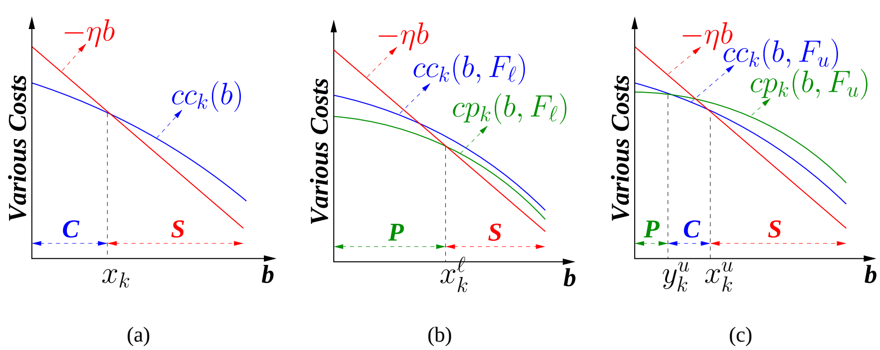


Fig. 2. Illustration of the threshold property. S, P and C in the figure represents the actions stop, probe and continue, respectively. (a) $\mathcal{S}_k$ is characterized by the threshold x_k. (b) and (c) depicts the stopping sets corresponding to the distributions F_ℓ and F_u, respectively. The distributions are such that $F_\ell \ge_{st} F_u$.

B. Optimal Probing Sets

Now, similar to the stopping set $\mathcal{S}_k^\ell$, one can define an *optimal probing set* $\mathcal{P}_k^\ell$ as, the set of all b such that at (b, F_ℓ) it is better to probe than to stop or continue, i.e.,

$$\mathcal{P}_k^\ell \quad := \quad \Big\{ b : cp_k(b, F_\ell) \le \min\{-\eta b, cc_k(b, F_\ell)\} \Big\}. \tag{15}$$

Alternatively, $\mathcal{P}_k^\ell$ is simply the difference of the sets $\mathcal{Q}_k^\ell$ and $\mathcal{S}_k^\ell$, i.e., $\mathcal{P}_k^\ell = \mathcal{Q}_k^\ell \setminus \mathcal{S}_k^\ell$. In our numerical work we observed that, similar to the stopping set, the probing set was characterized by an upper bound y_k^ℓ as illustrated in Fig. 2(b) (where $y_k^\ell = x_k^\ell$) and 2(c). At the time of this writing we have not proved such a result. However, we strongly believe that it is true and make the following conjecture,

Conjecture 1: For $1 \leq k \leq N-1$, for any F_ℓ, if $b_2 \in \mathcal{P}_k^\ell$ then for any $b_1 < b_2$ we have $b_1 \in \mathcal{P}_k^\ell$. ■

The above conjecture, along with Lemma 7 (in the next section, where we show that, for any F_ℓ, $\mathcal{S}_k \subseteq \mathcal{Q}_k^\ell$), is the reason for the illustration in Fig. 2(b) to not contain a region where it is optimal to continue. Whenever $x_k^\ell > x_k$, from Lemma 7 we know that it is optimal to probe for every $b \in [x_k, x_k^\ell)$. Then the above conjecture suggests that it is optimal to probe for every $b < x_k^\ell$, leaving no region where it is optimal to continue. Unlike x_k^ℓ, the thresholds y_k^ℓ are stage dependent. In fact, from our numerical work, we observe that y_k^ℓ are increasing with k.

C. Stopping Sets: Stage Independence Property

In this section we will prove that the stopping sets are identical across the stages. First we need the following results.

Lemma 6: Suppose $\mathcal{S}_k \subseteq \mathcal{Q}_k^\ell$ for some $k = 1, 2, \cdots, N-1$, and some F_ℓ, then for every $b \in \mathcal{S}_k$ we have $J_k(b, F_\ell) = J_N(b, F_\ell)$.

Proof: Fix a $b \in \mathcal{S}_k \subseteq \mathcal{Q}_k^\ell$. From the definition of the set $\mathcal{Q}_k^\ell$ (from (14)) we know that at (b, F_ℓ) it is optimal to either stop or probe. Further, after probing the state is $\max\{b, R_\ell\} \in \mathcal{S}_k$ (from Lemma 1) so that it is optimal to stop. Combining these observations we can write

$$\begin{aligned} J_k(b, F_\ell) &= \min\Big\{ -\eta b, \eta\delta + \mathbb{E}_\ell\Big[J_k(\max\{b, R_\ell\})\Big]\Big\} \\ &= \min\Big\{ -\eta b, \eta\delta - \eta\mathbb{E}_\ell\Big[\max\{b, R_\ell\}\Big]\Big\} \\ &= J_N(b, F_\ell). \end{aligned}$$

■

Lemma 7: For $1 \leq k \leq N-1$ and for any $F_\ell \in \mathcal{F}$ we have $\mathcal{S}_k \subseteq \mathcal{Q}_k^\ell$.

Proof Outline: For a formal proof see Appendix VIII-D. We only provide an outline here.

Step 1: First we show that (Lemma 10 in Appendix VIII-D), if there exists an F_u such that $F_\ell \geq_{st} F_u$ and for every $1 \leq k \leq N-1$, $\mathcal{S}_k \subseteq \mathcal{Q}_k^u$ then we have $\mathcal{S}_k \subseteq \mathcal{Q}_k^\ell$. This requires the total stochastic ordering of $\mathcal{F}$.

Step 2: Next we show that (Lemma 11 in Appendix VIII-D) the minimal distribution F_m satisfies, for every $1 \leq k \leq N-1$, $\mathcal{S}_k \subseteq \mathcal{Q}_k^m$.

The proof is complete by recalling that $F_\ell \geq_{st} F_m$ for every $F_\ell \in \mathcal{F}$ (Assumption 1) and then using, in *Step 1*, F_m in the place of F_u. ■

The following are the main theorems of this section,

Theorem 2: For $1 \leq k \leq N-2$, $\mathcal{S}_k = \mathcal{S}_{k+1}$.

Proof: From Lemma 4-(iv) we already know that $\mathcal{S}_k \subseteq \mathcal{S}_{k+1}$. Here, we will show that $\mathcal{S}_k \supseteq \mathcal{S}_{k+1}$. Fix a $b \in \mathcal{S}_{k+1} \subseteq \mathcal{S}_{k+2}$. From Lemma 7 we know that $\mathcal{S}_{k+1} \subseteq \mathcal{Q}_{k+1}^\ell$ and $\mathcal{S}_{k+2} \subseteq \mathcal{Q}_{k+2}^\ell$, for every F_ℓ. Then, applying

Lemma 6 we can write, $J_{k+2}(b, F_\ell) = J_{k+1}(b, F_\ell) = J_N(b, F_\ell)$. Thus,

$$\begin{aligned} cc_{k+1}(b) &= \tau + \mathbb{E}_A\Big[J_{k+2}(b, F_{L_{k+2}})\Big] \\ &= \tau + \mathbb{E}_A\Big[J_{k+1}(b, F_{L_{k+1}})\Big] \\ &= cc_k(b) \end{aligned}$$

Now, since $b \in \mathcal{S}_{k+1}$ we have $-\eta b \leq cc_{k+1}(b) = cc_k(b)$ which implies that $b \in \mathcal{S}_k$. ■

Remark: It is worth pointing out the parallels and the differences of the above result with that in [22, Section 4.4, pp-165], where it is shown that the *one-step-look ahead* rule is optimal, but for the case where the rewards are exactly revealed. There, as in our Lemma 6, the key idea is to show that the cost-to-go functions, at every stage k, are identical for every state within the stopping set. For our case, to apply Lemma 6, it was further essential for us to prove Lemma 7 showing that for every F_ℓ, $\mathcal{S}_k \subseteq \mathcal{C}_k^\ell$. Now, for $\delta = 0$, Lemma 7 trivially holds, since at (b, F_ℓ) it is always optimal to probe (so that $\mathcal{C}_k^\ell = \Re_+$). Further, $\delta = 0$ can be thought of as the case where the rewards are exactly revealed. Thus, Theorem 2 can be considered as a generalization of the result in [22, Section 4.4] for the case, $\delta > 0$. ■

For any F_ℓ, the stopping sets $\mathcal{S}_k^\ell$ are also stage independent.

Theorem 3: For $1 \leq k \leq N-2$ and for any F_ℓ, $\mathcal{S}_k^\ell = \mathcal{S}_{k+1}^\ell$.

Proof: Similar to the proof of Theorem 2, here we need to show that, for $b \in \mathcal{S}_k^\ell$, the probing and continuing costs satisfy, $cp_{k+1}(b, F_\ell) = cp_{k+2}(b, F_\ell)$ and $cc_{k+1}(b, F_\ell) = cc_{k+2}(b, F_\ell)$, respectively. See Appendix VIII-E. ■

From the previous subsection we already know that the optimal policy is completely characterized, at every stage k, in terms of the thresholds x_k and $\{x_k^\ell, y_k^\ell\}$ for every F_ℓ. From Theorem 2 and 3 in this section, we further know that the thresholds x_k and x_k^ℓ are stage independent and hence have to be computed only once for stage $N-1$, thus simplifying the overall computation of the optimal policy.

V. Optimal Policy within the Complete Class Π

In this section we will consider the complete class Π. Here we will only informally discuss the possible structure of the optimal policy within Π. Formal analysis is still in progress at the time of this writing. However, we will write the expressions for the recursive cost-to-go functions (analogous to the ones in (5) and (6)).

Recall that a policy within Π, at stage k, is in general allowed to take decisions based on the entire history. Formally, at stage k, let $b_k = \max\Big\{R_i : i \leq k, \text{ relay } i \text{ has been probed}\Big\}$ and $\mathcal{F}_k = \Big\{F_{L_i} : i \leq k, \text{ relay } i \text{ is unprobed}\Big\}$, then the decision at stage k is based on $(b_k, \mathcal{F}_k)$. Thus the set of all possible states (State Space SS_k) at stage k is

$$SS_k = \Big\{(b, \mathcal{G}) : b \in \Re_+, \mathcal{G} = \{G_1, G_2, \cdots, G_\ell\}, 0 \leq \ell \leq k, G_i \in \mathcal{F}, 1 \leq i \leq \ell\Big\}. \tag{16}$$

Again the actions available are stop, probe and continue. Further, if the action is to probe then one has to decide which relay to probe, among the several ones awake at stage k.

Let J_k, $k = N, N-1, \cdots, 1$, represent the optimal cost-to-go at stage k (for simplicity we are again using J_k), then, $J_N(b) = -\eta b$, and

$$J_N(b, \mathcal{G}) = \min\Big\{ -\eta b, \eta\delta + \min_{G_i \in \mathcal{G}} \mathbb{E}_{G_i}\Big[J_N(\max\{b, R_i\}, \mathcal{G} \setminus \{G_i\})\Big]\Big\}. \tag{17}$$

For stage $k = N-1, N-2, \cdots, 1$ we have

$$J_k(b) = \min\Big\{ -\eta b, \mathbb{E}_A\Big[J_k(b, \{F_{L_{k+1}}\})\Big]\Big\}, \tag{18}$$

$$J_k(b, \mathcal{G}) = \min\Big\{ -\eta b, \eta\delta + \min_{G_i \in \mathcal{G}} \mathbb{E}_{G_i}\Big[J_k(\max\{b, R_i\}, \mathcal{G} \setminus \{G_i\})\Big], \mathbb{E}_A\Big[J_{k+1}(b, \mathcal{G} \cup \{F_{L_{k+1}}\})\Big]\Big\}. \tag{19}$$

A. Discussion on the Last Stage N

Consider the scenario where the decision process enters the last stage N. Given the best reward value b, among the relays that have been probed, and the set $\mathcal{G}$ of reward distributions of the unprobed relays, the source has to decide whether to stop or probe a relay (note that there is no continue action available at the last stage). This decision problem is similar to the one studied by Chaporkar and Proutiere in [13] in the context of channel selection, which we briefly describe now. Given a set of channels with different channel gain distributions, a transmitter has to choose a channel for its transmissions. The transmitter can probe a channel to know its channel gain. Probing all the channels will enable the transmitter to select the best channel but at the cost of reduced effective transmission time within the coherence period. On the other hand, probing only a few channels may deprive the transmitter of the opportunity to transmit on a better channel. The transmitter is interested in *maximizing* its *throughput* within the coherence period, which is analogous to the *cost* (combination of delay and effective reward, see (3)) in our case, which we are trying to *minimize*.

The authors in [13], for their channel probing problem, prove that the one-step-look-ahead (OSLA) rule is optimal. Thus, given the channel gain of the best channel (among the channels probed so far) and a collection of channel gain distributions of the unprobed channels, it is optimal to stop and transmit on the best channel if and only if the throughput obtained by doing so is greater than the average throughput obtained by probing any unprobed channel and then stopping (i.e., transmitting on the new-best channel). Further they prove that if the set of channel gain distributions is totally stochastically ordered (see Assumption 1), then it is optimal to probe the channel whose distribution is stochastically largest among all the unprobed channels. Applying the result of [13] to our model we can conclude that OSLA is optimal once the decision process enters the last stage N. Thus given a state, $(b, \mathcal{G})$, at stage N it is optimal to stop if the cost of stopping is less than the cost of probing any distribution from $\mathcal{G}$ and then stopping. Otherwise it is optimal to probe the stochastically largest distribution from $\mathcal{G}$.

B. Discussion on Stages $k = N-1, N-2, \cdots, 1$

For the other stages $k = N-1, N-2, \cdots, 1$, one can begin by defining the stopping sets $\mathcal{S}_k$, $\mathcal{S}_k^{\mathcal{G}}$ and stopping/probing set $\mathcal{Q}_k^{\mathcal{G}}$ analogous to the ones in (12), (13) and (14). Note that this time we need to define $\mathcal{S}_k^{\mathcal{G}}$ and $\mathcal{Q}_k^{\mathcal{G}}$ for a set of distributions $\mathcal{G}$ unlike in the earlier case where we had defined these sets only for a given distribution F_ℓ. We conjecture that it is possible to show the results analogous to the ones in Section IV, namely

Theorem 2 and 3 where we prove that the stopping sets are identical for every stage k. Formal analysis is still in progress at the time of writing this paper. However, in the numerical results section (Section VI) while performing value iteration we observed that our conjecture is true, at least for the example considered for the numericals.

Finally, from the discussion in the previous sub-section we know that, at stage N, suppose it is optimal to probe when the state is $(b, \mathcal{G})$ then it is best to probe the stochastically largest distribution from $\mathcal{G}$. We also conjecture that such a result will hold for every stage k, which is true for the numerical example considered in Section VI.

VI. Numerical Results

The optimal policy within the restricted class (Section III and IV) is allowed to keep only one unprobed relay awake in addition to the best probed relay, while within the complete class (Section V), the optimal policy can keep all the unprobed relays awake. We will refer to the former policy as RST-OPT (to be read as, ReSTricted-OPTimal) and the latter as simply, GLB-OPT (read as, GLoBal-OPTimal). In this section we will compare the performance of the RST-OPT against GLB-OPT. First we will briefly describe the relay selection example considered for the numerical work.

Recall the local forwarding problem described in Section II. The source and sink are seperated by a distance of $v_0 = 10$ unit (see Fig. 1). The radius of the communication region is set to 1 unit. There are $N = 5$ relays within the forwarding region $\mathcal{L}$. These are uniformly located within $\mathcal{L}$. To enable us to perform value iteration (i.e., recursively solve the Bellman equation to obtain optimal value and the optimal policy), we discretize the forwarding region $\mathcal{L}$ by considering a set of 20 equally spaced points within $\mathcal{L}$ and then rounding the location of each relay to the respective closest point. Recall the reward expression from (1). We have fixed, $\Gamma N_0 = 1$, $\beta = 2$ and $a = 0.5$. We finally normalize the reward to take values within the interval $[0, 1]$ and then quantize it to one of the 100 equally spaced points within $[0, 1]$. The inter-wake-up times $\{U_k\}$ are exponential with mean $\tau = 0.2$.

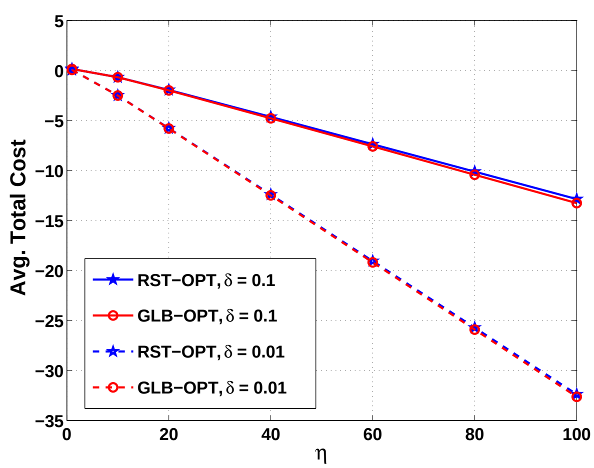
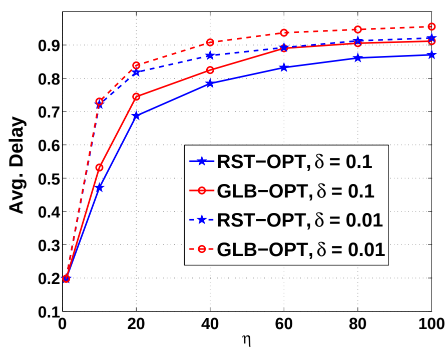


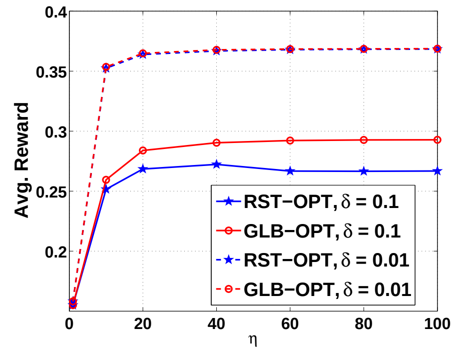
Fig. 3. Average total cost vs. the Lagrange multiplier η.

In Fig. 3 we plot the average total cost (see (3)), incurred by RST-OPT and GLB-OPT, as a function of the

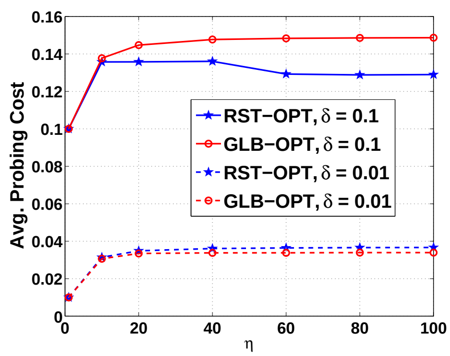
Lagrange multiplier η for two different values of the probing cost δ (namely $\delta = 0.1$ and $\delta = 0.01$). The total cost is decreasing with η. First, observe that the total cost incurred, by either policy, for $\delta = 0.01$ is smaller than for $\delta = 0.1$. This is because, when the probing cost is smaller, each of the policy will end up probing more relays, thus accruing a larger reward and yielding a lower cost. Next, since GLB-OPT is optimal over a larger class of policies, we know that, for a given δ, GLB-OPT should incur a smaller cost than RST-OPT. However, interestingly from the plot we observe that the difference between both the costs is small. Also, from the figure, note that the cost difference for $\delta = 0.01$ is smaller than that for $\delta = 0.1$. This is because, as the probing cost is decreased, both the policies will start behaving identically by probing most of the relays until they stop. Finally, when there is no cost for probing, i.e., for $\delta = 0$, we expect both the policies to be identical.



(a)



(b)



(c)

Fig. 4. Individual components of the total cost in Fig. 3 as functions of η. (a) Avg. Delay (b) Avg. Reward and (c) Avg. Probing Cost.

In Fig. 4(a), 4(b), and 4(c) we plot the individual components (namely delay, reward and probing cost, respectively) of the total cost, as a function of η. All the curves in these figures are increasing with η, except for $\delta = 0.1$ where the average reward and probing cost of RST-OPT shows a small decrease when η varies from 40 to 60. As η decreases we observe, from Fig. 4(a), that all the average delay curves converge to 0.2 (recall that $\tau = 0.2$ is the average time until the first relay wakes up). This is because, for small values of η, the source values the delay more (recall (3)) and hence always forwards to the relay that wakes up first, irrespective of its reward. For the

same reason the average probing cost curves in Fig. 4(c) converge to their respective δ value which is the cost for probing a single relay. Similarly the average reward in Fig. 4(b) converges to the average reward value of the first relay.

Finally on the computational complexity of both the policies. To obtain the policy GLB-OPT we had to recursively solve the Bellman equations in (18) and (19), for every stage k and every possible state at stage k, starting from the last stage N, see (17). The total number of all possible states at stage k, i.e., the cardinality of the state space SS_k in (16), grows exponentially with the cardinality of $\mathcal{F}$ (assuming that $\mathcal{F}$ is discrete like in our numerical example). It also grows exponentially with the stage index k. While in contrast for computing RST-OPT, since within the restricted class, at any time, only one unprobed relay is kept awake, the state space size grows linearly with the cardinality of $\mathcal{F}$. Further, the size of the state space does not grow with the number of relays N. From our analysis in Section IV we know that the stopping sets are threshold based and moreover the thresholds (x_k and $\{x_k^\ell : F_\ell \in \mathcal{F}\}$) are stage independent, so that these thresholds have to be computed only once (namely for stage $N-1$), further reducing the complexity of RST-OPT.

RST-OPT can also be regarded as *energy efficient* in the sense that it keeps only one unprobed relay (in addition to the best probed one) awake while instructing the other unprobed ones to switch to a low power OFF state (i.e., sleep state). While GLB-OPT operates by keeping all the relays, that have woken up thus far, awake. The close to optimal performance of RST-OPT, its computational simplicity and energy efficiency is motivating us to apply RST-OPT at each hop enroute to the sink in a large sensor network and study its end-to-end performance, which is a part of our ongoing work.

VII. Conclusion and Future Scope

We considered the sequential decision problem of choosing a relay node for the next hop, by a forwarding node, in the presence of random sleep-wake cycling. In the model, we have incorporated the energy cost of probing a relay to learn its channel gain. We have analysed a restricted class of policies where any policy is allowed to keep at most one unprobed relay awake, in addition to the best probed relay. The optimal policy for the restricted class (RST-OPT) was characterized in terms of, the stopping sets and the stopping/probing sets. First, we showed that the stopping sets are threshold in nature. Further, we proved that the thresholds, characterizing the stopping sets, are stage independent, thus simplifying the computation of RST-OPT in comparison with the global optimal, GLB-OPT (whose decisions at any stage is based on the entire history). Numerical work confirmed that the performance of RST-OPT is close to that of GLB-OPT.

As mentioned earlier, a part of our ongoing work is to study the end-to-end performance (i.e., total delay and overall power consumption) of RST-OPT for a given network. Another interesting direction is to apply cooperative techniques for forwarding in sleep-wake cycling networks and study its benefits.

REFERENCES

[1] J. Kim, X. Lin, and N. Shroff, "Optimal Anycast Technique for Delay-Sensitive Energy-Constrained Asynchronous Sensor Networks," in *INFOCOM 2009. The 28th Conference on Computer Communications. IEEE*, April 2009, pp. 612–620.

[2] Q. Cao, T. Abdelzaher, T. He, and J. Stankovic, "Towards Optimal Sleep Scheduling in Sensor Networks for Rare-Event Detection," in *IPSN '05: Proceedings of the 4th international symposium on Information processing in sensor networks*, 2005.

[3] K. Premkumar, A. Kumar, and J. Kuri, "Distributed Detection and Localization of Events in Large Ad Hoc Wireless Sensor Networks," in *Communication, Control, and Computing, 2009. Allerton 2009. 47th Annual Allerton Conference on*, sept. 2009, pp. 178 –185.

[4] K. P. Naveen and A. Kumar, "Tunable Locally-Optimal Geographical Forwarding in Wireless Sensor Networks with Sleep-Wake Cycling Nodes," in *INFOCOM 2010. The 29th Conference on Computer Communications. IEEE*, march 2010, pp. 1–9.

[5] S. Y. Ni, Y. C. Tseng, Y. S. Chen, and J. P. Sheu, "The Broadcast Storm Problem in a Mobile Ad hoc Network," in *Proceedings of the 5th annual ACM/IEEE international conference on Mobile computing and networking*, ser. MobiCom '99. New York, NY, USA: ACM, 1999, pp. 151–162.

[6] F. Ye, G. Zhong, S. Lu, and L. Zhang, "Gradient broadcast: a robust data delivery protocol for large scale sensor networks," *Wirel. Netw.*, vol. 11, pp. 285–298, May 2005.

[7] D. P. Bertsekas and J. N. Tsitsiklis, "An Analysis of Stochastic Shortest Path Problems," *Mathematics of Operations Research*, vol. 16, pp. 580–595, 1991.

[8] K. P. Naveen and A. Kumar, "Relay Selection with Partial Information in Wireless Sensor Networks," Dept. of Electrical Communication Engg., IISc, Bangalore, Tech. Rep. ARXIV Report Number 1101.3835, 2011. [Online]. Available: http://arxiv.org/abs/1101.3835

[9] K. Akkaya and M. Younis, "A Survey on Routing Protocols for Wireless Sensor Networks," *Ad Hoc Networks*, vol. 3, pp. 325–349, 2005.

[10] M. Zorzi and R. R. Rao, "Geographic Random Forwarding (GeRaF) for Ad Hoc and Sensor Networks: Multihop Performance," *IEEE Transactions on Mobile Computing*, vol. 2, pp. 337–348, 2003.

[11] J. Xu, B. Peric, and B. Vojcic, "Performance of Energy-Aware and Link-Adaptive Routing Metrics for Ultra Wideband Sensor Networks," *Mob. Netw. Appl.*, vol. 11, no. 4, pp. 509–519, 2006.

[12] M. Rossi, M. Zorzi, and R. R. Rao, "Statistically Assisted Routing Algorithms (SARA) for Hop Count Based Forwarding in Wireless Sensor Networks," *Wirel. Netw.*, vol. 14, no. 1, pp. 55–70, 2008.

[13] P. Chaporkar and A. Proutiere, "Optimal Joint Probing and Transmission Strategy for Maximizing Throughput in Wireless Systems," *Selected Areas in Communications, IEEE Journal on*, vol. 26, no. 8, pp. 1546–1555, October 2008.

[14] N. B. Chang and M. Liu, "Optimal Channel Probing and Transmission Scheduling for Opportunistic Spectrum Access," in *MobiCom '07: Proceedings of the 13th annual ACM international conference on Mobile computing and networking*, 2007, pp. 27–38.

[15] T. S. Ferguson, *Optimal Stopping and Applications*. Dept. of Mathematics, UCLA. [Online]. Available: http://www.math.ucla.edu/~tom/Stopping/Contents.html

[16] M. Sakaguchi, "Dynamic Programming of Some Sequential Sampling Design," *Journal of Mathematical Analysis and Applications*, vol. 2, no. 3, pp. 446 – 466, 1961.

[17] S. Karlin, *Stochastic Models and Optimal Policy for Selling an Asset, Studies in Applied Probability and Management Science / edited by Kenneth J. Arrow, Samuel Karlin, Herbert Scarf*. Stanford University Press, Stanford, Calif, 1962.

[18] S. Ikuta, "Optimal Stopping Problem with Uncertain Recall," *Journal of the Operations Research Society of Japan 31*, p. 145170, 1988.

[19] D. B. Rosenfield, R. D. Shapiro, and D. A. Butler, "Optimal Strategies for Selling an Asset," *Management Science*, vol. 29, no. 9, pp. pp. 1051–1061, 1983.

[20] W. Stadje, "An Optimal Stopping Problem with Two Levels of Incomplete Information," *Mathematical Methods of Operations Research*, vol. 45, pp. 119–131, 1997.

[21] P. Thejaswi, J. Zhang, M. O. Pun, H. Poor, and D. Zheng, "Distributed Opportunistic Scheduling with Two-Level Probing," *Networking, IEEE/ACM Transactions on*, vol. 18, no. 5, pp. 1464 –1477, oct. 2010.

[22] D. P. Bertsekas, *Dynamic Programming and Optimal Control, Vol. I*. Athena Scientific, 2005.

[23] J. Kim, X. Lin, and N. B. Shroff, "Optimal Anycast Technique for Delay Sensitive Energy-Constrained Asynchronous Sensor Networks," 2008, Technical Report, Purdue University. [Online]. Available: http://web.ics.purdue.edu/~kim309/Kim08tech3.pdf

VIII. APPENDICES

A. Motivation for the Particular Reward Structure

In this section we will motivate the reason for considering the particular reward structure (namely a weighed ratio of progress and power, see (1)). We were interested in minimizing end-to-end delay subject to a constraint on the total transmission power. We considered solving the local forwarding problem for the proposed reward in (1). However we assumed a *deterministic channel model* (i.e., no fading or equivalently $|H_\ell|^2 = 1$ in (1)) so that the exact reward value is revealed by the relays upon waking up. Therefore our earlier work [4] can be applied to obtain a simple threshold (on reward) based rule. We have applied this policy at each hop towards the sink in a network comprising of 500 nodes that are randomly sleep-wake cycling. We observed that, for $a = 0.5$, the performance of our policy (in terms of total average delay and average transmission power) is comparable to the performance of the LOCAL-OPT algorithm (a Bellman-Ford based algorithm) proposed by Kim et al. [23]. The performance tradeoff curves are shown in Fig. 5. Though these observations were for the model with deterministic channels, we want to consider the same reward structure in the present work as well, where the channel gains are random and hence the source has to probe a relay to learn its reward.

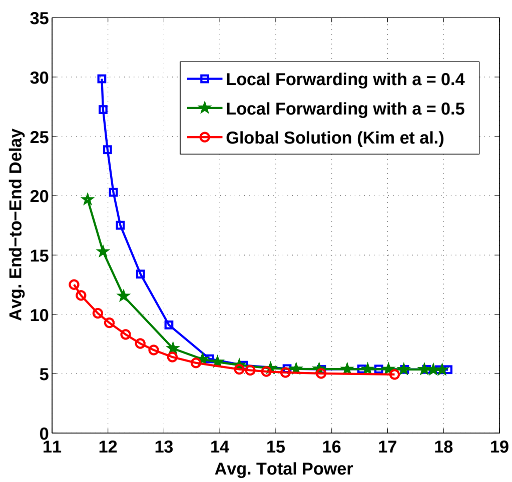


Fig. 5. Curves showing the tradeoff between the average end-to-end delay and the total power for various policies.

B. Proof of Lemma 2

First, let us make a note of the following lemma,

Lemma 8: If $F_\ell \geq_{st} F_u$, then for any b we have $F_{\max\{b,R_\ell\}} \geq_{st} F_{\max\{b,R_u\}}$, where the distributions of R_ℓ and R_u are F_ℓ and F_u, respectively.

Proof: The proof easily follows by noting that

$$F_{\max\{b,R_\ell\}}(x) = \begin{cases} 0 & \text{if } x < b \\ F_\ell(x) & \text{otherwise.} \end{cases}$$

■

Proof of Lemma 2-(i): Proof is by induction. For stage N we know that $J_N(b) = -\eta b$, and hence is decreasing in b. Then, from (4) it is clear that $J_N(b, F_\ell)$ is also decreasing in b. Thus, the monotonicity properties holds for $k = N$. Suppose $J_{k+1}(b)$ and $J_{k+1}(b, F_\ell)$ are decreasing in b for some $k+1 = 2, 3, \cdots, N$. We will show that the result holds for k as well.

Recall the expressions of $J_k(b)$ and $J_k(b, F_\ell)$ ((10) and (11) respectively): $J_k(b) = \min\Big\{-\eta b, cc_k(b)\Big\}$ and $J_k(b, F_\ell) = \min\Big\{-\eta b, cp_k(b, F_\ell), cc_k(b, F_\ell)\Big\}$. Thus to complete the proof it is sufficient to show that $cc_k(b)$, $cp_k(b, F_\ell)$ and $cc_k(b, F_\ell)$ are decreasing in b.

From the induction hypothesis, it is easy to see that $cc_k(b)$ (in (7)) is decreasing in b, using which we can conclude that $J_k(b)$ (see (10)) is decreasing in b. Now that we have establised $J_k(b)$ is decreasing in b, it will immediately follow (from (9)) that the probing cost $cp_k(b, F_\ell)$ is also decreasing in b.

Again using the induction argument observe that $\min\Big\{J_{k+1}(b, F_\ell), J_{k+1}(b, F_{L_{k+1}})\Big\}$ is decreasing in b so that the continuing cost $cc_k(b, F_\ell)$ (in (8)) is also decreasing. ■

Proof of Lemma 2-(ii): Consider stage N and recall the optimal cost-to-go function $J_N(b, F_\ell)$ from (4),

$$J_N(b, F_\ell) = \min\Big\{-\eta b, \eta\delta - \eta\mathbb{E}_\ell\Big[\max\{b, R_\ell\}\Big]\Big\}.$$

Using the definition of stochastic ordering (Definition 1) and Lemma 8 we can write

$$\mathbb{E}_\ell\Big[\max\{b, R_\ell\}\Big] \geq \mathbb{E}_u\Big[\max\{b, R_u\}\Big],$$

so that the result holds for stage N. Suppose the result holds for some $k+1 = 2, 3, \cdots, N$. We will show that the cost of probing and the cost of continuing, in (9) and (8), both satisfy $cp_k(b, F_\ell) \leq cp_k(b, F_u)$ and $cc_k(b, F_\ell) \leq cc_k(b, F_u)$. Then the result easily follows for stage k by recalling (from (11)) that $J_k(b, F_\ell) = \min\Big\{-\eta b, cp_k(b, F_\ell), cc_k(b, F_\ell)\Big\}$.

From Lemma 2-(i) we know that $J_k(b)$ is decreasing in b. Again using Definition 1 and Lemma 8, we can conclude that

$$\mathbb{E}_\ell\Big[J_k(\max\{b, R_\ell\})\Big] \leq \mathbb{E}_u\Big[J_k(\max\{b, R_u\})\Big]$$

so that $cp_k(b, F_\ell) \leq cp_k(b, F_u)$ (see (9)).

From the induction argument we know that $J_{k+1}(b, F_\ell) \leq J_{k+1}(b, F_u)$ so that

$$\min\Big\{J_{k+1}(b, F_\ell), J_{k+1}(b, F_{L_{k+1}})\Big\} \leq \min\Big\{J_{k+1}(b, F_u), J_{k+1}(b, F_{L_{k+1}})\Big\}.$$

Therefore $cc_k(b, F_\ell) \leq cc_k(b, F_u)$ (see (8)). ■

Proof of Lemma 2(iii): This result is very intuitive, since with more number of stages to go, one is expected to accrue less average cost. However, we prove it here for completeness.

Again the proof is by induction. For stage $N-1$ we know that, $J_{N-1}(b) = \min\Big\{-\eta b, cc_k(b)\Big\}$. Replacing $-\eta b$ with $J_N(b)$ in the previous expression we obtain $J_{N-1}(b) \leq J_N(b)$. Next, consider a state of the form (b, F_ℓ). The

cost of probing $cp_{N-1}(b, F_\ell)$ can be bounded as follows,

$$\begin{aligned} cp_{N-1}(b, F_\ell) &= \eta\delta + \mathbb{E}_\ell\Big[J_{N-1}(\max\{b, R_\ell\})\Big] \\ &\leq \eta\delta + \mathbb{E}_\ell\Big[J_N(\max\{b, R_\ell\})\Big] \\ &= \eta\delta - \eta\mathbb{E}_\ell\Big[\max\{b, R_\ell\}\Big] \end{aligned}$$

where, in the second inequality we have used, $J_{N-1}(b) \leq J_N(b)$, which we have already shown. Therefore,

$$\begin{aligned} J_{N-1}(b, F_\ell) &= \min\Big\{-\eta b, c^p_{N-1}(b, F_\ell), c^c_{N-1}(b, F_\ell)\Big\} \\ &\leq \min\Big\{-\eta b, c^p_{N-1}(b, F_\ell)\Big\} \\ &\leq \min\Big\{-\eta b, \eta\delta - \eta\mathbb{E}_\ell\Big[\max\{b, R_\ell\}\Big]\Big\} \\ &= J_N(b, F_\ell). \end{aligned}$$

Thus we have shown the result for stage $N-1$. Suppose the result holds for some stage $k+1 = 2, 3, \cdots, N-1$. i.e., $J_{k+1}(b) \leq J_{k+2}(b)$ and $J_{k+1}(b, F_\ell) \leq J_{k+2}(b, F_\ell)$. Using the induction hypothesis, the cost of continuing $cc_k(b)$ can be bounded as

$$\begin{aligned} cc_k(b) &= \tau + \mathbb{E}_A\Big[J_{k+1}(b, F_{k+1})\Big] \\ &\leq \tau + \mathbb{E}_A\Big[J_{k+2}(b, F_{k+2})\Big] \\ &= c^c_{k+1}(b). \end{aligned}$$

Then $J_k(b) \leq J_{k+1}(b)$ (see (10)). Recall the expression of $J_k(b, F_\ell)$ from (11). To show that $J_k(b, F_\ell) \leq J_{k+1}(b, F_\ell)$, it is sufficient to show that $cp_k(b, F_\ell) \leq cp_{k+1}(b, F_\ell)$ and $cc_k(b, F_\ell) \leq cc_{k+1}(b, F_\ell)$. First consider the probing cost,

$$\begin{aligned} cp_k(b, F_\ell) &= \eta\delta + \mathbb{E}_\ell\Big[J_k(\max\{b, R_\ell\})\Big] \\ &\leq \eta\delta + \mathbb{E}_\ell\Big[J_{k+1}(\max\{b, R_\ell\})\Big] \\ &= cp_{k+1}(b, F_\ell) \end{aligned}$$

where, in the second inequality we have used $J_k(b) \leq J_{k+1}(b)$, which we have already shown. Finally, consider the cost of continuing,

$$\begin{aligned} cc_k(b, F_\ell) &= \tau + \mathbb{E}_A\Big[\min\{J_{k+1}(b, F_\ell), J_{k+1}(b, F_{L_{k+1}})\}\Big] \\ &\leq \tau + \mathbb{E}_A\Big[\min\{J_{k+2}(b, F_\ell), J_{k+2}(b, F_{L_{k+2}})\}\Big] \\ &= cc_{k+1}(b, F_\ell), \end{aligned}$$

where the second inequality is obtained by simply using the induction argument. ■

C. *Proof of Lemma 5*

The following property about the $\min$ operator will be useful while proving Lemma 5.

Lemma 9: If $x_1, x_2, \cdots, x_j$ and $y_1, y_2, \cdots, y_j$ in $\Re$, are such that, $x_i - y_i \leq x_1 - y_1$ for all $i = 1, 2, \cdots, j$, then

$$\min\{x_1, x_2, \cdots, x_j\} - \min\{y_1, y_2, \cdots, y_j\} \leq x_1 - x_2 \tag{20}$$

Proof: Suppose $\min\{y_1, y_2, \cdots, y_j\} = y_i$, for some $1 \leq i \leq j$, then the LHS of (20) can be written as,

$$LHS = \min\{x_1, x_2, \cdots, x_j\} - y_i \leq x_i - y_i.$$

The proof is complete by recalling that we are given, $x_i - y_i \leq x_1 - y_1$. ■

Proof of Lemma 5: Since $J_N(b)$ is simply $-\eta b$ we have, for stage N, $J_N(b_1) - J_N(b_2) = \eta(b_2 - b_1)$. Also, for a given distribution F_ℓ and for $b_2 > b_1$,

$$\begin{aligned} cp_N(b_1, F_\ell) - cp_N(b_2, F_\ell) &= \eta \mathbb{E}_\ell \Big[\max\{b_2, R_\ell\} - \max\{b_1, R_\ell\} \Big] \\ &\leq \eta(b_2 - b_1), \end{aligned}$$

which, along with Lemma 9, implies that $J_N(b_1, F_\ell) - J_N(b_2, F_\ell) \leq \eta(b_2 - b_1)$. Suppose for some stage $k+1 = 1, 2, \cdots, N$ we have $J_{k+1}(b_1) - J_{k+1}(b_2) \leq \eta(b_2 - b_1)$ and $J_{k+1}(b_1, F_\ell) - J_{k+1}(b_2, F_\ell) \leq \eta(b_2 - b_1)$. Then we will show that all the inequalities listed in the lemma will hold for stage k as well. First, a simple application of the induction hypothesis will yield,

$$\begin{aligned} cc_k(b_1) - cc_k(b_2) &= \mathbb{E}_A \Big[J_{k+1}(b_1, F_{L_k}) - J_{k+1}(b_2, F_{L_k}) \Big] \\ &\leq \eta(b_2 - b_1), \end{aligned}$$

which in turn (along with Lemma 9) implies $J_k(b_1) - J_k(b_2) \leq \eta(b_2 - b_1)$, using which we can write

$$\begin{aligned} cp_k(b_1, F_\ell) - cp_k(b_2, F_\ell) &= \mathbb{E}_\ell \Big[J_k(\max\{b_1, R_\ell\}) - J_k(\max\{b_2, R_\ell\}) \Big] \\ &\leq \mathbb{E}_\ell \Big[\eta(\max\{b_2, R_\ell\} - \max\{b_1, R_\ell\}) \Big] \\ &\leq \eta(b_2 - b_1). \end{aligned}$$

Define $\mathcal{L}_\ell$ as the set of all distributions that are stochastically greater than ℓ, i.e., $\mathcal{L}_\ell := \{F_t \in \mathcal{F} : F_t \geq_{st} F_\ell\}$. Let $\mathcal{L}_\ell^c := \mathcal{F} \setminus \mathcal{L}_\ell$. From Assumption 1 (that $\mathcal{F}$ is totally stochastically ordered), it follows that $\mathcal{L}_\ell^c$ contains all distributions in $\mathcal{F}$ which are stochastically smaller than F_ℓ. Then, using Lemma 2(ii) the cost of continuing can be simplified as,

$$\begin{aligned} cc_k(b_1, F_\ell) - cc_k(b_2, F_\ell) &= \int_{\mathcal{L}_\ell} (J_{k+1}(b_1, F_t) - J_{k+1}(b_2, F_t)) dA(t) \\ &\quad + \int_{\mathcal{L}_\ell^c} (J_{k+1}(b_1, F_\ell) - J_{k+1}(b_2, F_\ell)) dA(t). \end{aligned}$$

Now, a straight forward application of the induction argument gives the desired result. Finally, the induction argument is complete, by using Lemma 9, to conclude that $J_k(b_1, F_\ell) - J_k(b_2, F_\ell) \leq \eta(b_2 - b_1)$. ■

D. Proof of Lemma 7

As discussed in the outline of the proof of Lemma 7, the result immediately follows once we prove *Step 1* and *Step 2*.

Lemma 10: Suppose F_u is a distribution such that for all $k = 1, 2, \cdots, N-1$, $\mathcal{S}_k \subseteq \mathcal{Q}_k^u$. Then for any distribution $F_\ell \geq_{st} F_u$ we have $\mathcal{S}_k \subseteq \mathcal{Q}_k^\ell$.

Proof: We will first show that $\mathcal{S}_{N-1} \subseteq \mathcal{C}_{N-1}^\ell$. Fix a $b \in \mathcal{S}_{N-1}$. Then $b \in \mathcal{Q}_{N-1}^u$ (because it is given that $\mathcal{S}_{N-1} \subseteq \mathcal{Q}_{N-1}^u$), so that using the definition of the set $\mathcal{Q}_{N-1}^u$ (from (14)) we can write

$$\min\Big\{-\eta b, \eta\delta + \mathbb{E}_u\Big[J_{N-1}(\max\{b, R_u\})\Big]\Big\} \leq cc_{N-1}(b, F_u). \tag{21}$$

Since $b \in \mathcal{S}_{N-1}$, for any distribution F_s, the minimum of the cost of stopping and the cost of probing can be simplified as

$$\begin{aligned} \min\Big\{-\eta b, \eta\delta + \mathbb{E}_s\Big[J_{N-1}(\max\{b, R_s\})\Big]\Big\} &= \min\Big\{-\eta b, \eta\delta - \eta\mathbb{E}_s\Big[\max\{b, R_s\}\Big]\Big\} \\ &= J_N(b, F_s) \end{aligned} \tag{22}$$

where, in the first equality we have replaced $J_{N-1}(\max\{b, R_s\})$ by the cost of stopping $(-\eta\max\{b, R_s\})$. This is because, after probing we are still at stage k with the new state being $\max\{b, R_s\}$, which is also in $\mathcal{S}_{N-1}$ (Lemma 1) and in $\mathcal{S}_{N-1}$ we know that it is optimal to stop, so that $J_{N-1}(\max\{b, R_s\}) = -\eta\max\{b, R_s\}$.

Thus we are given, $J_N(b, F_u) \leq cc_{N-1}(b, F_u)$ (use (22) in (21)). Also from Lemma 2-(ii), since we are given $F_\ell \geq_{st} F_u$, we have $J_N(b, F_\ell) \leq J_N(b, F_u)$. Combining these we can write

$$J_N(b, F_\ell) \leq J_N(b, F_u) \leq cc_{N-1}(b, F_u). \tag{23}$$

By recalling (22) and using (21), we need to show, $J_N(b, F_\ell) \leq cc_{N-1}(b, F_\ell)$ to conclude that $b \in \mathcal{Q}_{N-1}^\ell$.

Now for any distribution $F_s \in \mathcal{F}$ define $\mathcal{L}_s := \{t \in \mathcal{L} : F_t \geq_{st} F_s\}$ i.e., $\mathcal{L}_s$ is the set of all distributions in $\mathcal{F}$ that are stochastically greater than F_s. Let $\mathcal{L}_s^c = \mathcal{L} \setminus \mathcal{L}_s$. Because of the condition imposed on $\mathcal{F}$ that, it is totally stochastically ordered (Assumption 1), $\mathcal{L}_s^c$ contains all distributions in $\mathcal{F}$ that are stochastically smaller than F_s. Further, for $F_\ell \geq_{st} F_u$ we have $\mathcal{L}_\ell \subseteq \mathcal{L}_u$. Then recalling the expression for $cc_{N-1}(b, F_u)$ from (8) we can write

$$\begin{aligned} cc_{N-1}(b, F_u) &= \tau + \mathbb{E}_A\Big[\min\{J_N(b, F_u), J_N(b, F_{L_N})\}\Big] \\ &= \tau + \int_{\mathcal{L}_u} J_N(b, F_t)\ dA(t) + \int_{\mathcal{L}_u^c} J_N(b, F_u)\ dA(t) \\ &= \tau + \int_{\mathcal{L}_\ell} J_N(b, F_t)\ dA(t) + \int_{\mathcal{L}_u \setminus \mathcal{L}_\ell} J_N(b, F_t)\ dA(t) + \int_{\mathcal{L}_u^c} J_N(b, F_u)\ dA(t), \end{aligned}$$

where to obtain second equality we have used Lemma 2-(ii) and the definition of $\mathcal{L}_u$. For any $F_t \in \mathcal{L}_u \setminus \mathcal{L}_\ell$ we know that $F_t \geq_{st} F_u$ so that $J_N(b, F_t) \leq J_N(b, F_u)$ (again from Lemma 2(ii)). Thus replacing $J_N(b, F_t)$ by $J_N(b, F_u)$ in the middle integral above we obtain

$$cc_{N-1}(b, F_u) \quad \leq \quad \tau + \int_{\mathcal{L}_\ell} J_N(b, F_t)\ dA(t) + \int_{\mathcal{L}_\ell^c} dA(t)\ J_N(b, F_u) \tag{24}$$

From (23) and (24) we see that we have an inequality of the following form

$$J_N(b,F_\ell) \le J_N(b,F_u) \le c + pJ_N(b,F_u),$$

where $c = \tau + \int_{\mathcal{L}_\ell} J_N(b,F_t)\ dA(t)$ and $p = \int_{\mathcal{L}_\ell^c} dA(t)$. Since $p \in [0,1]$ we can write

$$J_N(b,F_\ell)(1-p) \le J_N(b,F_u)(1-p)$$

rearranging which, we obtain

$$\begin{aligned} J_N(b,F_\ell) &\le pJ_N(b,F_\ell) + J_N(b,F_u) - pJ_N(b,F_u) \\ &\le pJ_N(b,F_\ell) + c + pJ_N(b,F_u) - pJ_N(b,F_u) \\ &= c + pJ_N(b,F_\ell). \end{aligned}$$

The proof for stage $N-1$ is complete by noting that $c + pJ_N(b,F_\ell)$ is $cc_{N-1}(b,F_\ell)$.

Suppose that for some $k+1 = 2,3,\cdots,N-1$ we have $\mathcal{S}_{k+1} \subseteq \mathcal{Q}_{k+1}^\ell$. We will have to show that the same holds for stage k. Fix any $b \in \mathcal{S}_k$, then for any distribution F_s, exactly as in (22)

$$\begin{aligned} \min\Big\{-\eta b, \eta\delta + \mathbb{E}_s\Big[J_k(\max\{b,R_s\})\Big]\Big\} &= \min\Big\{-\eta b, \eta\delta - \eta\mathbb{E}_s\Big[\max\{b,R_s\}\Big]\Big\} \\ &= J_N(b,F_s). \end{aligned} \tag{25}$$

Thus the hypothesis $\mathcal{S}_k \subseteq \mathcal{Q}_k^u$ implies $J_N(b,F_u) \le cc_k(b,F_u)$ and we need to show $\mathcal{S}_k \subseteq \mathcal{Q}_k^\ell$, i.e., $J_N(b,F_\ell) \le cc_k(b,F_\ell)$. Proceeding as before (see (24)) we obtain

$$cc_k(b,F_u) \le \tau + \int_{\mathcal{L}_\ell} J_{k+1}(b,F_t)\ dA(t) + \int_{\mathcal{L}_\ell^c} dA(t)\ J_{k+1}(b,F_u).$$

Now using Lemma 6, we conclude

$$cc_k(b,F_u) \le \tau + \int_{\mathcal{L}_\ell} J_{k+1}(b,F_t)\ dA(t) + \int_{\mathcal{L}_\ell^c} dA(t)\ J_N(b,F_u).$$

Note that the conditions required to apply Lemma 6 hold i.e., $b \in \mathcal{S}_{k+1}$ (since $\mathcal{S}_k \subseteq \mathcal{S}_{k+1}$ from Lemma 4-(iv)) and $\mathcal{S}_{k+1} \subseteq \mathcal{Q}_{k+1}^u$ (this is given).

Thus, again we have an inequality of the form $J_N(b,F_\ell) \le J_N(b,F_u) \le c' + pJ_N(b,F_u)$ (where $c' = \tau + \int_{\mathcal{L}_\ell} J_{k+1}(b,F_t)\ dA(t)$). As before we can show that $J_N(b,F_\ell) \le c' + pJ_N(b,F_\ell)$. Finally the proof is complete by showing that $c' + pJ_N(b,F_\ell) = cc_k(b,F_\ell)$ as follows,

$$\begin{aligned} cc_k(b,F_\ell) &= \tau + \int_{\mathcal{L}_\ell} J_{k+1}(b,F_t)\ dA(t) + \int_{\mathcal{L}_\ell^c} J_{k+1}(b,F_\ell)\ dA(t) \\ &= c' + pJ_N(b,F_\ell), \end{aligned} \tag{26}$$

where to replace $J_{k+1}(b,F_\ell)$ by $J_N(b,F_\ell)$ we have to again apply Lemma 6. However this time $\mathcal{S}_{k+1} \subseteq \mathcal{Q}_{k+1}^\ell$, is by the induction hypothesis. ∎

To use the above lemma we still require a distribution F_u satisfying $\mathcal{S}_k \subseteq \mathcal{Q}_k^u$, for every k. The minimal distribution F_m turns out to be useful in this context.

Lemma 11: For every $k = 1, 2, \cdots, N-1$, $\mathcal{S}_k \subseteq \mathcal{Q}_k^m$.

Proof: Since F_m is a minimal distribution (see Assumption 1), from Lemma 2-(ii) we know that

$$J_{k+1}(b, F_{L_{k+1}}) \leq J_{k+1}(b, F_m).$$

Using the above expression in (8) and then recalling (7), we can write $cc_k(b, F_m) = cc_k(b)$. Finally, the result follows from the definition of the sets $\mathcal{Q}_k^m$ and $\mathcal{S}_k$. ■

E. Proof of Theorem 3

Proof: Recalling the definition of the set $\mathcal{S}_k^\ell$ (from (13)), for any $b \in \mathcal{S}_{k+1}^\ell$ we have

$$-\eta b \leq \min\Big\{cp_{k+1}(b, F_\ell), cc_{k+1}(b, F_\ell)\Big\}.$$

Suppose, as in Theorem 2, we can show that for any $b \in \mathcal{S}_k^\ell$, the various costs at stages k and $k+1$ are same, i.e., $cp_k(b, F_\ell) = cp_{k+1}(b, F_\ell)$ and $cc_k(b, F_\ell) = cc_{k+1}(b, F_\ell)$, then this would imply, $\mathcal{S}_k^\ell \supseteq \mathcal{S}_{k+1}^\ell$. The proof is complete by recalling that we already have $\mathcal{S}_k^\ell \subseteq \mathcal{S}_{k+1}^\ell$ (from Lemma 4-(iv)).

Fix a $b \in \mathcal{S}_{k+1}^\ell$. To show that $cp_k(b, F_\ell) = cp_{k+1}(b, F_\ell)$, first using Lemma 4-(ii) and Theorem 2, note that $\mathcal{S}_{k+1}^\ell \subseteq \mathcal{S}_{k+1} = \mathcal{S}_k$. Since $b \in \mathcal{S}_{k+1}$ the cost of probing is

$$\begin{aligned} cp_{k+1}(b, F_\ell) &= \eta\delta + \mathbb{E}_\ell\Big[J_{k+1}(\max\{b, R_\ell\})\Big] \\ &= \eta\delta - \eta\mathbb{E}_\ell\Big[\max\{b, R_\ell\}\Big] \end{aligned}$$

where, to obtain the second equality, note that $\max\{b, R_\ell\} \in \mathcal{S}_k$ (Lemma 1) and hence at $\max\{b, R_\ell\}$ it is optimal to stop, so that $J_{k+1}(\max\{b, R_\ell\}) = -\eta \max\{b, R_\ell\}$. Similarly, since b is also in $\mathcal{S}_k$ the cost of probing at stage k, $cp_k(b, F_\ell)$, is again $\eta\delta - \eta\mathbb{E}_\ell\Big[\max\{b, R_\ell\}\Big]$.

Finally, following the same procedure used in Theorem 2 to show $cc_k(b) = cc_{k+1}(b)$, one can prove that $cc_k(b, F_\ell) = cc_{k+1}(b, F_\ell)$. ■